

A Frequentist Approach to Computer Model Calibration

Raymond K. W. Wong^{*1}, Curtis B. Storlie^{†2}, and Thomas C. M. Lee^{‡3}

¹Department of Statistics, Iowa State University

²Los Alamos National Laboratory

³Department of Statistics, University of California, Davis

September 04, 2015

Abstract

This paper considers the computer model calibration problem and provides a general frequentist solution. Under the proposed framework, the data model is semi-parametric with a nonparametric discrepancy function which accounts for any discrepancy between the physical reality and the computer model. In an attempt to solve a fundamentally important (but often ignored) identifiability issue between the computer model parameters and the discrepancy function, this paper proposes a new and identifiable parametrization of the calibration problem. It also develops a two-step procedure for estimating all the relevant quantities under the new parameterization. This estimation procedure is shown to enjoy excellent rates of convergence and can be straightforwardly implemented with existing software. For uncertainty quantification, bootstrapping is adopted to construct confidence regions for the quantities of interest. The practical performance of the proposed methodology is illustrated through simulation examples and an application to a computational fluid dynamics model.

Keywords: bootstrap; inverse problem; model misspecification; semi-parametric modeling; surrogate model; uncertainty analysis

^{*}Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A. Email: raywong@iastate.edu

[†]Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A. Email: storlie@lanl.gov

[‡]Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, U.S.A. Email: tcmlee@ucdavis.edu

1 Introduction

In many scientific studies, complex mathematical models, implemented as computer codes, are often used to model the physical reality (see, e.g., Santner *et al.*, 2003; Fang *et al.*, 2010). Such computer codes are also known as *computer models*, and can only be executed when certain model parameters are pre-specified. The goal of computer model *calibration* is to find the model parameter values that allow the computer model to best reproduce the physical reality.

In the computer model calibration problem (Kennedy and O’Hagan, 2001), an output y is observed from the physical reality ζ at n locations of a p -variate input $\mathbf{x} = (x_1, \dots, x_p)^\top$:

$$y_i = \zeta(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

where ε_i is the measurement error for the i -th observation. It is assumed that the user can select the values of the design locations $\mathbf{x}_1, \dots, \mathbf{x}_n$. A computer model $\eta(\mathbf{x}, \boldsymbol{\theta})$, also called the *simulator*, can be used to approximate the physical reality $\zeta(\mathbf{x})$ when its model parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^\top$ is selected to be close to the unknown ideal value $\boldsymbol{\theta}_0$. To account for the discrepancy between the simulator and the physical reality, one can introduce a discrepancy function $\delta_0(\mathbf{x})$ and the model of the experimental data $\{\mathbf{x}_i, y_i\}_{i=1}^n$ becomes

$$y_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}_0) + \delta_0(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n. \quad (1)$$

To use (1) in practice, one first needs to estimate $\boldsymbol{\theta}_0$ and δ_0 , which requires evaluation of $\eta(\mathbf{x}, \boldsymbol{\theta})$ at many different values of \mathbf{x} and $\boldsymbol{\theta}$. However, the evaluation of $\eta(\mathbf{x}, \boldsymbol{\theta})$ is often very computationally expensive due to the complex nature of the mathematical models. This complication can be alleviated via the use of a surrogate model, often referred to as an *emulator*, for the simulator $\eta(\mathbf{x}, \boldsymbol{\theta})$ (e.g., Currin *et al.*, 1991; Kennedy and O’Hagan, 2001; Higdon *et al.*, 2004; Drignei and Morris, 2006; Conti *et al.*, 2009; Reich *et al.*, 2009). Typically a Gaussian process (GP) is assumed for the emulator to allow for a flexible model of the simulator. Moreover, the discrepancy function δ_0 is also often modeled by a GP. In order to construct an emulator, an additional set of outputs

y'_i 's generated from the simulator is obtained at m design locations $(\mathbf{x}'_1, \boldsymbol{\theta}'_1), \dots, (\mathbf{x}'_m, \boldsymbol{\theta}'_m)$. Thus, there are two observed data sets: the *experimental data* set $\{\mathbf{x}_i, y_i\}_{i=1}^n$ obtained from the physical reality and the *simulator data* set $\{\mathbf{x}'_i, \boldsymbol{\theta}'_i, y'_i\}_{i=1}^m$ generated by the simulator $\eta(\mathbf{x}, \boldsymbol{\theta})$. Note that the experimental and simulator data sets are fundamentally different. The simulator data are generated solely from the simulator and so they are not directly related to the physical reality. Moreover, the designs of the two data sets do not have to match. That is, $\{\mathbf{x}'_i\}$ does not necessarily equal to $\{\mathbf{x}_i\}$ either in value or sampling distribution. In this setting, the goal is to estimate $\boldsymbol{\theta}_0$, δ_0 and η .

Traditionally, this estimation problem is solved within a Bayesian framework (e.g., Kennedy and O'Hagan, 2001; Higdon *et al.*, 2004; Oakley and O'Hagan, 2004; Bayarri *et al.*, 2007; Higdon *et al.*, 2008; Storlie *et al.*, 2014), partly because of the GP's ability to incorporate uncertainty about the surrogate model for $\eta(\mathbf{x}, \boldsymbol{\theta})$, and a similar ability to provide uncertainty about the discrepancy function δ_0 . However, in this paper we are interested in solving this problem from a purely frequentist perspective with a nonparametric discrepancy function, and at the same time also accounting for uncertainty in the model parameters, the surrogate model, and the discrepancy function. To the best of our knowledge, there has been no previous attempt to solve the computer model calibration problem in this manner. Yet, there are several reasons for doing this:

1. The proposed approach is conceptually clean and simple, easy to understand, and can be paired with any choice of surrogate model.
2. It delivers a complementary calibration result to the Bayesian approach, allowing each approach to provide some qualitative and quantitative confirmation of the other.
3. Researchers using computational models may be opposed to the Bayesian calibration approach because of the complex prior assumptions of GP and of an identifiability issue between the surrogate model and the discrepancy function (see Section 2.1). Consequently they may prefer the proposed frequentist approach instead.

The proposed frequentist approach explicitly accounts for all potentially important sources of uncertainty, and is a viable alternative to the Bayesian approach. While any statistical model makes assumptions, there are fewer assumptions necessary in the proposed approach than in the Bayesian

counterpart. For example, prior distributions on the surrogate model and the discrepancy function are replaced with “smoothness” assumptions and an emphasis is intuitively placed on the ability to predict the experimental data via cross-validation. Empirical evidence shows that the proposed frequentist approach tends to outperform existing Bayesian methods on the examples considered.

Lastly we remark that there are some frequentist solutions for the computer model calibration problem. The most common one involves obtaining the maximum likelihood estimate (MLE) for θ directly by evaluating η sequentially in an optimization routine (e.g., Vecchia and Cooley, 1987; Jones *et al.*, 1998; Huang *et al.*, 2006). If η is computationally expensive then as before a surrogate model could be used in place of η for the purpose of obtaining an MLE for θ . However, this latter approach must be used with caution as to not ignore the estimation uncertainty in the surrogate model for η , which can often be substantial. Also, unlike the proposed approach, most previous frequentist methods suffer from the same shortcoming: no discrepancy modeling is included in their formulation (i.e., $\delta_0 = 0$). While some surrogate models may be very good approximations, no model is perfect and neglecting the discrepancy can be a major pitfall (Bayarri *et al.*, 2007; Brynjarsdóttir and O’Hagan, 2014). One notable exception is the pioneering work by Joseph and Melkote (2009), which substantially improves previous frequentist model calibration methods by assuming a parametric form for the discrepancy function δ_0 . Our work brings frequentist calibration approaches to the next level in several respects: (i) we allow for a nonparametric (or semi-parametric) δ_0 to provide a practical representation of ignorance for the discrepancy; (ii) we present a rigorous justification for the proposed estimates, and (iii) we deliver a simple, yet general procedure to account for the uncertainty in the surrogate model, the model parameters, and the discrepancy. As discussed in detail below, the use of a nonparametric δ_0 gives rise to an identifiability issue and some theoretical challenges, which separates our work from others.

The rest of this article is organized as follows. In Section 2 we introduce an identifiable definition of the calibration problem, and provide a practical estimation procedure. Section 3 illustrates how the bootstrap methodology can be applied to provide uncertainty quantification for parameters of interest. Theoretical backup for our estimation procedure is presented in Section 4. The practical performance of our methodology is illustrated via a real life computational fluid dynamics model in

Section 5, while concluding remarks are given in Section 6. Additional material including simulation experiments and technical details are provided in a separate online supplemental document.

2 The Proposed Approach

Consider the semi-parametric model (1) for the physical reality. Despite its popularity under the Bayesian framework, this model is not identifiable in the frequentist regime, where $\boldsymbol{\theta}_0$ and δ_0 are treated as fixed but unknown quantities to be estimated. In the following we will first discuss this non-identifiability issue and provide intuitive and identifiable definitions for $\boldsymbol{\theta}_0$ and δ_0 under model (1) (Sections 2.1). We then develop an efficient method for estimating these quantities when the simulator η is known (Section 2.2) and unknown (Section 2.3). As to be seen below, the proposed frequentist framework is very general and covers many practical situations. Its estimation procedure can be paired with any existing optimization techniques and nonparametric regression methods, therefore provides an effective and flexible approach to estimate $\boldsymbol{\theta}_0$ and δ_0 with convenient implementation.

2.1 An Identifiable Formulation

In model (1), the discrepancy function δ_0 is assumed to be an unconstrained smooth function, and as such will be estimated using nonparametric regression techniques. To see the non-identifiability of (1), consider two different values $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ for $\boldsymbol{\theta}$, and write $\delta_1(\mathbf{x}) = \zeta(\mathbf{x}) - \eta(\mathbf{x}, \boldsymbol{\theta}_1)$ and $\delta_2(\mathbf{x}) = \zeta(\mathbf{x}) - \eta(\mathbf{x}, \boldsymbol{\theta}_2)$. As both $(\boldsymbol{\theta}_1, \delta_1)$ and $(\boldsymbol{\theta}_2, \delta_2)$ give the same distribution for y , model (1) in general is unidentifiable. In the Bayesian paradigm, with the help of suitable priors, the posterior distributions for $\boldsymbol{\theta}_0$ and δ_0 are technically well-defined. However, this unidentifiability still poses many problems at a more foundational level. For example, it forbids any meaningful construction of uncertainty measures for $\boldsymbol{\theta}_0$ or δ_0 , as these target quantities do not have unique definitions. We also note that previous frequentist methods bypass this issue by setting $\delta_0 = 0$.

Now we provide natural and sensible definitions of $\boldsymbol{\theta}_0$ and δ_0 to achieve identifiable modeling. Write the spaces of model parameters and inputs as $\boldsymbol{\Theta}$ and \mathcal{X} , respectively. We propose the

following model for the physical reality ζ :

$$\zeta(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}_0) + \delta_0(\mathbf{x}),$$

where η is the simulator and δ_0 is the discrepancy function, both to be modeled by smooth functions.

We define

$$\boldsymbol{\theta}_0 = \operatorname{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_{\mathcal{X}} \{\zeta(\mathbf{x}) - \eta(\mathbf{x}, \boldsymbol{\theta})\}^2 dF(\mathbf{x}), \quad (2)$$

where the distribution F characterizes a weighting scheme of \mathbf{x} . For simplicity, we assume that F is the sampling distribution of \mathbf{x} . This is mostly sensible since the sampling design, being user controllable, would characterize one's attention to ζ over different values of \mathbf{x} . Thus this should also be reflected in the definition of $\boldsymbol{\theta}_0$. For common sampling design schemes, F is intuitively defined as an identity function and thus $\boldsymbol{\theta}_0$ is the minimizer of the integrated squared error between ζ and η . However, if a weighting scheme, other than the sampling distribution of \mathbf{x} , is intended, our procedure can be modified easily by introducing a simple reweighting (to M_n defined below in (4)) to estimate the corresponding $\boldsymbol{\theta}_0$.

Under a mild regularity assumption (Assumption 5 below), the solution of the minimization problem in (2) is unique, and hence it is straightforward to see that both $\boldsymbol{\theta}_0$ and δ_0 are identifiable.

The definition (2) for $\boldsymbol{\theta}_0$ is logical and natural, as it aligns with the intuition that $\boldsymbol{\theta}_0$ should be the value that makes η closest to ζ , and δ_0 is used to account for the remainder. Using other arguments or motivation, it is of course possible to provide alternative identifiable definitions for $\boldsymbol{\theta}_0$ and δ_0 . For example, Brynjarsdóttir and O'Hagan (2014) argue that the the "best fitting" model is not always the most desirable. However, our definition (2) leads directly to a straightforward and easily implementable estimation procedure, as to be described below. We also note that our definition is in the same spirit as Walker (2013).

2.2 Estimation when the Simulator is Known

Suppose we observe the output y of the physical reality ζ at n locations $\mathbf{x}_1, \dots, \mathbf{x}_n$; i.e., $y_i = \zeta(\mathbf{x}_i) + \varepsilon_i$, $i = 1, \dots, n$, where ε_i is the i -th observation error. These errors are assumed to be

independent and have mean 0. For simplicity, we assume the p -variate \mathbf{x}_i 's have been scaled such that $\mathcal{X} = [0, 1]^p$. With the above modeling of ζ , the observations are assumed to follow

$$y_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}_0) + \delta_0(\mathbf{x}_i) + \varepsilon_i. \quad (3)$$

For the estimation of $\boldsymbol{\theta}_0$ and δ_0 , we first consider a simpler situation for which η can be assumed known or evaluated rapidly. For cases when this assumption is not true, we will estimate η with a second set of samples and the details are described later in Section 2.3.

The definitions of $\boldsymbol{\theta}_0$ and δ_0 naturally motivate a two-step estimation procedure:

1. *Estimation of $\boldsymbol{\theta}_0$* : compute the estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ as the solution to the following minimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} M_n(\boldsymbol{\theta}) \quad \text{with} \quad M_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \{y_i - \eta(\mathbf{x}_i, \boldsymbol{\theta})\}^2. \quad (4)$$

2. *Estimation of δ_0* : estimate δ_0 by applying any nonparametric regression method to the “data” $\{\mathbf{x}_i, y_i - \eta(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\}_{i=1}^n$.

This estimation procedure has its beauty in flexibility and ease of implementation. It can be coupled with any (global) minimization technique in Step 1 and any nonparametric regression method in Step 2. For example, for convenience one could adopt an existing and fast off-the-shelf minimization routine for Step 1, and wavelet technique for Step 2 if one believes δ_0 is mostly smooth with a few sharp jumps. Also, as the estimation of $\boldsymbol{\theta}_0$ and that of δ_0 are separated, there is no need to re-run the minimization in Step 1 when choosing the smoothing parameter in the nonparametric regression step. Thus in general this estimation procedure can be made very fast computationally with suitable choices of minimization and nonparametric regression techniques. For all the numerical illustrations in this paper, we adopt the genetic optimization using derivative (Sekhon and Mebane, 1998) for Step 1 and smoothing spline ANOVA (Wahba, 1990) for Step 2. The theoretical support for using these estimators are provided in Section 4 under a fixed design setting.

2.3 Estimation when the Simulator is Approximated with an Emulator

This subsection handles the situation when the simulator η is unknown or expensive to run. As mentioned in introduction, a common strategy to overcome this issue is to approximate η with a surrogate model, also known as an emulator. The emulator is estimated nonparametrically from a second set of observations obtained by running the simulator for different combinations of inputs \mathbf{x} and model parameters $\boldsymbol{\theta}$. Write the simulator output at m design locations $(\mathbf{x}'_1, \boldsymbol{\theta}'_1), \dots, (\mathbf{x}'_m, \boldsymbol{\theta}'_m)$ as $\mathbf{y}' = (y'_1, \dots, y'_m)^\top$. They are assumed to follow

$$y'_j = \eta(\mathbf{x}'_j, \boldsymbol{\theta}'_j) + \tau_j, \quad j = 1, \dots, m,$$

where τ_j 's are independent random errors with mean zero. The underlying physical model is typically continuous and in theory these τ_j 's should not be needed. However, they are included here to allow for numerical jitter in the simulator evaluations due to various reasons such as convergence criteria.

The proposed approach proceeds as follows. We first use $\{\mathbf{x}'_j, \boldsymbol{\theta}'_j, y'_j\}_{j=1}^m$ to fit an emulator via a nonparametric regression method such as SS-ANOVA (Wahba, 1990). Denote the resulting emulator as $\hat{\eta}$. We then treat $\hat{\eta}$ as fixed and replace η by $\hat{\eta}$ in (3) when estimating $\boldsymbol{\theta}_0$ and δ_0 via the estimation procedure proposed in Section 2.2. We note that the parameter $\boldsymbol{\theta}$ can be constrained to a particular domain, as is often done (albeit with the ability to be more informative) in the Bayesian calibration approach via a prior distribution.

In the above description the estimation of η and δ_0 is done separately. One could alternatively perform a joint estimation by combining the two estimation problems into one semi-parametric optimization problem. However, in this case (as in the Bayesian approach) the experimental data $\{\mathbf{x}_i, y_i\}_{i=1}^n$ will influence the estimation of the emulator. This could be beneficial by providing a smaller variance to the emulator, but it could also be problematic by inserting a large bias into the emulator toward the reality function. The more conservative approach taken here eliminates this potential bias by removing the influence of the experimental data on the emulator.

To the best of our knowledge, this approach to obtain a point estimate for the calibration

problem with discrepancy ($\boldsymbol{\theta}_0$ and δ_0) for a computationally demanding simulator η has not been attempted until now. The above description does not yet account for the uncertainty in the estimation of the emulator, model parameters, or the discrepancy function. However, this issue can easily be addressed via the bootstrap method (see, e.g., Efron and Tibshirani, 1993), to be described next.

3 Uncertainty Quantification using Bootstrap

In computer modeling, bootstrap has been applied successfully to quantify the uncertainty in the emulator for the purpose of sensitivity analysis (SA) and uncertainty analysis (UA) for computationally demanding simulators (e.g., Storlie *et al.*, 2009, 2013). It is therefore expected that bootstrap will also provide equally successful results for the current calibration problem. However, it is noted that this calibration problem is far more complicated than SA and UA due to the additional estimation of $\boldsymbol{\theta}_0$ and δ_0 .

Let the point estimates of the unknown parameters in model (1) be obtained as described above and denoted as $\hat{\boldsymbol{\theta}}$, $\hat{\eta}$, and $\hat{\delta}$. These define an estimate for the data generating process for both the simulator data \mathbf{y}' and experimental data \mathbf{y} . A bootstrap sample for the calibration problem can be generated with the following steps:

1. (Optional) Re-sample the designs in both data sets if the data were generated under random designs.
2. Produce B bootstrap samples by re-sampling centered residuals.
3. Re-estimate the parameters to obtain B bootstrap estimates of $\boldsymbol{\theta}$, η , and δ . Denote them as $\hat{\boldsymbol{\theta}}_b^*$, $\hat{\eta}_b^*$, and $\hat{\delta}_b^*$, $b = 1, \dots, B$, respectively.

The resulting bootstrap sample of the estimates can be used to obtain a bootstrap confidence region for most quantities of interest.

In calibration problems, confidence intervals for the elements of $\boldsymbol{\theta}_0$ and pointwise confidence bands for δ_0 are usually of interest. For example, to obtain a confidence interval for the first element $\theta_{0,1}$ of $\boldsymbol{\theta}_0$, one can find the $\alpha/2$ and $(1 - \alpha/2)$ sample quantiles from $\{\hat{\theta}_{1,1}^*, \dots, \hat{\theta}_{B,1}^*\}$, where

$\hat{\theta}_{b,1}^*$ represents the first element of $\hat{\boldsymbol{\theta}}_b^*$ for $b = 1, \dots, B$. Denote these quantiles as $z_{\alpha/2}^*$ and $z_{1-\alpha/2}^*$, respectively. The required confidence interval is then given by $(z_{\alpha/2}^*, z_{1-\alpha/2}^*)$. A confidence interval for a prediction of the physical reality ζ at any new input \mathbf{x}_{new} and the pointwise confidence band for δ_0 can be obtained in a similar fashion.

Since our estimation procedure involves nonparametric regression, the impact of bias may lead to incorrect asymptotic coverage of the aforementioned bootstrap confidence regions (see, e.g., Härdle and Bowman, 1988; Hall, 1992a,b). In the literature, there are two common strategies for correcting the coverage: undersmoothing and oversmoothing. As shown in Hall (1992a), undersmoothing is a simpler and more effective strategy than oversmoothing. Our estimation procedure can be easily modified to incorporate undersmoothing; e.g., by choosing a smaller smoothing parameter. However, the gain in practical performance of these strategies are usually small and most of these strategies involve another ad-hoc choice of the amount of under- or over-smoothing. Moreover, it is not uncommon to ignore this bias issue, essentially resulting in the use of non-adjusted confidence regions as described above; see, e.g., Efron and Tibshirani (1993) and Ruppert *et al.* (2003). To keep the approach simple and adaptable to a wide class of nonparametric regression methods, we recommend using the non-adjusted confidence regions.

4 Theoretical Results

This section provides theoretical support to the proposed estimation procedure presented in Section 2. First recall that the estimation of η depends on a second independent sample generated from the simulator of size m . In practice this sample is typically much larger than the sample obtained from the physical reality; i.e., $m \gg n$. Thus, it is reasonable to assume that m approaches infinity at a faster rate than n in the asymptotic framework. If $m \rightarrow \infty$ fast enough, the asymptotics of $\hat{\boldsymbol{\theta}}$ and $\hat{\delta}$ would be similar to those under known η . Therefore, for simplicity and to speed up the development, in the following we assume η is known and derive the asymptotic properties of $\hat{\boldsymbol{\theta}}$ and $\hat{\delta}$ defined in Section 2.2.

Write $\hat{\boldsymbol{\theta}}$ and $\hat{\delta}$ as $\hat{\boldsymbol{\theta}}_n$ and $\hat{\delta}_n$ respectively to address their dependence on n . In the following, we assume that $\mathbf{x}_1, \dots, \mathbf{x}_n$ are fixed and use F_n to denote their empirical distribution

function. In addition, $\|\cdot\|_n$, $\|\cdot\|$ and $\|\cdot\|_E$ represent the $L_2(F_n)$ -norm, the $L_2(F)$ -norm and the Euclidean norm respectively. For two functions g and h , let $\langle g, h \rangle_n = \sum_{i=1}^n g(\mathbf{x}_i)h(\mathbf{x}_i)$ and $\langle g, h \rangle = \int_{\mathcal{X}} g(\mathbf{x})h(\mathbf{x})dF(\mathbf{x})$. With slight notation abuse, we also write $\langle y, g \rangle_n = (1/n) \sum_{i=1}^n y_i g(\mathbf{x}_i)$ and $\langle \varepsilon, g \rangle_n = (1/n) \sum_{i=1}^n \varepsilon_i g(\mathbf{x}_i)$. Lastly, write $g_{\boldsymbol{\theta}}(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta})$, $\mathcal{G} = \{g_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ and $\mathcal{G} - g = \{g_{\boldsymbol{\theta}} - g : \boldsymbol{\theta} \in \Theta\}$ for a function g . And, use $g_{\boldsymbol{\theta}}^{(j)}$ to represent the j -th order derivative of $g_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$ for $j = 1, 2$.

When deriving asymptotic results for similar statistical problems, it is relatively common to assume an independent and identically distributed (i.i.d.) random design, as it is easier than a fixed design to work with. However, for most practical calibration problems, the design is either fixed or correlated (e.g., Latin Hypercube sampling). Therefore the following results are developed under a fixed design, despite it is a more challenging setting than the i.i.d. random design. Note that model (3) is a semi-parametric model and we first approach the parametric part and establish the \sqrt{n} -consistency of $\hat{\boldsymbol{\theta}}_n$ in (see Theorem 1), where the difficulty lies in the existence of the discrepancy function δ_0 . The effect is similar to a regression model with misspecification.

As for the nonparametric part, δ_0 , we adopt the framework of Section 10.1 of van de Geer (2000) for penalized least squares estimation. We extend Theorem 10.2 of van de Geer (2000) to obtain the asymptotic behavior of $\hat{\delta}_n$ (see Lemma 2 of the supplemental document and Theorem 2), taking into account the effect of estimation error of $\hat{\boldsymbol{\theta}}_n$. Let the class of functions to which δ_0 belongs be \mathcal{H} . We suppose that \mathcal{H} is a cone. Under van de Geer's framework, the general form of the estimate of δ_0 is

$$\hat{\delta}_n = \underset{\delta \in \mathcal{H}}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^n \{y_i - g_{\hat{\boldsymbol{\theta}}_n}(\mathbf{x}_i) - \delta(\mathbf{x}_i)\}^2 + \lambda_n^2 J^v(\delta) \right], \quad (5)$$

where $v > 0$, $\lambda_n > 0$, $J : \mathcal{H} \rightarrow [0, \infty)$ is a pseudo-norm on \mathcal{H} . The λ_n is known as the smoothing parameter.

As an illustration, we provide the convergence rate of $\hat{\delta}_n$ for $p = 1$ if a penalized smoothing spline is used (see Corollary 1). This requires an additional orthogonality argument for the application of Theorem 2. We will write \mathbf{x} as x when $p = 1$.

Below are the assumptions needed for our theoretical results.

Assumption 1 (Error structure). $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) = \sigma^2$ for $i = 1, \dots, n$. Also, $\varepsilon_1, \dots, \varepsilon_n$ are uniformly sub-Gaussian; that is, there exist K and σ_0 such that

$$\max_{i=1, \dots, n} K^2 \{ \mathbb{E} \exp(\varepsilon_i^2 / K^2) - 1 \} \leq \sigma_0^2.$$

Assumption 2 (Parameter space). Θ is a totally bounded d -dimensional Euclidean space. That is, there exists an $R_0 > 0$ such that $\|\boldsymbol{\theta}\|_E \leq R_0$ for all $\boldsymbol{\theta} \in \Theta$.

Assumption 3 (Function class \mathcal{G}).

(a) There exists a $c_0 > 0$ such that $\|g_{\boldsymbol{\theta}} - g_{\boldsymbol{\theta}'}\|_n \leq c_0 \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_E$ for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$.

(b) $g_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. $g_{\boldsymbol{\theta}}^{(1)}(\mathbf{x})$ and $g_{\boldsymbol{\theta}}^{(2)}(\mathbf{x})$ are continuous with respect to \mathbf{x} over this neighborhood.

(c) $\sup_{\mathbf{x} \in \mathcal{X}} |g_{\boldsymbol{\theta}}^{(1)}(\mathbf{x})|$ and $\sup_{\mathbf{x} \in \mathcal{X}} |g_{\boldsymbol{\theta}}^{(2)}(\mathbf{x})|$ are bounded uniformly over a neighborhood of $\boldsymbol{\theta}_0$.

Assumption 4 (Convergence of design).

(a) $\sup_{\boldsymbol{\theta} \in \Theta} |\|\zeta - g_{\boldsymbol{\theta}}\|_n^2 - \|\zeta - g_{\boldsymbol{\theta}}\|^2| = o(1)$.

(b) Elements of $A_n - A$ are $o(1)$, where

$$A_n = (1/n) \sum_{i=1}^n \{g_{\boldsymbol{\theta}_0}^{(1)}(\mathbf{x}_i) g_{\boldsymbol{\theta}_0}^{(1)}(\mathbf{x}_i)^\top - \delta_0(\mathbf{x}_i) g_{\boldsymbol{\theta}_0}^{(2)}(\mathbf{x}_i)\},$$

$$A = \int_{\mathcal{X}} g_{\boldsymbol{\theta}_0}^{(1)}(\mathbf{x}) g_{\boldsymbol{\theta}_0}^{(1)}(\mathbf{x})^\top - \delta_0(\mathbf{x}) g_{\boldsymbol{\theta}_0}^{(2)}(\mathbf{x}) dF(\mathbf{x}),$$

are the second derivative of $M_n(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_0$ and that of $M(\boldsymbol{\theta})$ respectively.

(c) Euclidean norm of the first derivative of M_n evaluated at $\boldsymbol{\theta}_0$, $(1/n) \sum_{i=1}^n \delta_0(\mathbf{x}_i) g_{\boldsymbol{\theta}_0}^{(1)}(\mathbf{x}_i)$, is $\mathcal{O}(n^{-1/2})$.

Assumption 5 (Identification). A is strictly positive definite.

Assumption 6 (Discrepancy function).

(a) δ_0 is bounded.

(b) $J(\delta_0) < \infty$.

(c) There exist $K_1 > 0$ and $0 < \alpha < 2$ such that $H(u, \tilde{\mathcal{H}}, F_n) \leq K_1 u^{-\alpha}$, for all $u > 0$ and $n \geq 1$, where

$$\tilde{\mathcal{H}} = \left\{ \frac{\delta - \delta_0}{J(\delta) + J(\delta_0)} : \delta \in \mathcal{H}, J(\delta) + J(\delta_0) > 0 \right\}$$

and $H(u, \tilde{\mathcal{H}}, F_n)$ is the u -entropy of $\tilde{\mathcal{H}}$ for the $L_2(F_n)$ -metric (Definition 2.1 of van de Geer, 2000).

(d) $\sup_{\gamma \in \tilde{\mathcal{H}}} \|\gamma\|_n < \infty$.

The two main theorems and a corollary now follow. Their proofs can be found in Section S3 of the supplemental document.

Theorem 1 (Rates of convergence of $\hat{\boldsymbol{\theta}}_n$ and $g_{\hat{\boldsymbol{\theta}}_n}$). *Suppose that Assumptions 1-5 and 6(a) hold. We have $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_E = \mathcal{O}_p(n^{-1/2})$ and $\|g_{\hat{\boldsymbol{\theta}}_n} - g_{\boldsymbol{\theta}_0}\|_n = \mathcal{O}_p(n^{-1/2})$.*

Theorem 2 (Rate of convergence of $\hat{\delta}_n$). *Suppose that Assumptions 1-6 hold. Moreover, assume that $v > (2\alpha)/(2 + \alpha)$.*

(i) *If $J(\delta_0) > 0$ and $\lambda_n \asymp n^{-1/(2+\alpha)}$, we have $\|\hat{\delta}_n - \delta_0\|_n = \mathcal{O}_p(n^{-1/(2+\alpha)})$.*

(ii) *If $J(\delta_0) = 0$ and $J(\delta) > 0$ for all $\delta \in \mathcal{H} \setminus \{\delta_0\}$, we have*

$$\|\hat{\delta}_n - \delta_0\|_n = \mathcal{O}_p\left(\max\left\{n^{-1/2}, \lambda_n^{-2\alpha/(2v-2\alpha+v\alpha)} n^{-v/(2v-2\alpha+v\alpha)}\right\}\right).$$

Corollary 1 (Penalized smoothing spline). *Assume $p = 1$, $m \in \{1, 2, \dots\}$, $\mathcal{H} = \{\delta : [0, 1] \rightarrow \mathbb{R}, \int_0^1 \{\delta^{(m)}(x)\}^2 dx < \infty\}$ and $J(\delta) = [\int_0^1 \{\delta^{(m)}(x)\}^2 dx]^{1/2}$. And $\hat{\delta}_n$ is given in (5) with $v = 2$. Suppose the conditions of Theorem 1 hold and $J(\delta_0) < \infty$. Let $\boldsymbol{\psi} = (\psi_1, \dots, \psi_m)^\top$ where $\psi_k(x) = x^{k-1}$ for $k = 1, \dots, m$. Further, assume that the smallest eigenvalue of $\int \boldsymbol{\psi} \boldsymbol{\psi}^\top dF_n$ is bounded away from 0.*

(i) *If $J(\delta_0) > 0$, $\lambda_n \asymp n^{-m/(2m+1)}$, we have $\|\hat{\delta}_n - \delta_0\|_n = \mathcal{O}_p(n^{-m/(2m+1)})$.*

(ii) If $J(\delta_0) = 0$, we have

$$\|\hat{\delta}_n - \delta_0\|_n = \mathcal{O}_p\left(n^{-1/2} \max\left\{1, \lambda_n^{-1/(2m)}\right\}\right).$$

5 Application to a Computational Fluid Dynamics Model

In this section, the proposed approach is applied to a computational fluid dynamics (CFD) model of a bubbling fluidized bed (Lane *et al.*, 2014). The experimental apparatus used to produce the field data used here is the Carbon Capture Unit (C2U) housed at National Energy Technology Laboratory. The C2U unit is a bench-top carbon capture system, designed to mimic a post-combustion capture device that could be applied at a coal fired power plant. Section S1 of the supplemental document provides an illustration of the C2U system. A gas mixture (i.e., flue gas) flows through the bottom of the adsorber pictured on the right of the figure shown in Section S1 of the supplemental document and into the bed of solid sorbent, resulting in fluidization of the sorbent. At low temperature ($\sim 40^\circ \text{C}$), CO_2 will chemically bond to the solid sorbent and be effectively lifted out of the gas mixture. The solid sorbent would then circulate out of the adsorber, be stripped of CO_2 , and flow back into the adsorber. However, in this example, the goal was to isolate the fluid dynamics of the bubbling fluidized bed, and thus there is only nitrogen gas flowing through the solid sorbent material in the bed (i.e., no CO_2 adsorption is taking place). These data were collected as part of Department of Energy’s (DOE’s) Carbon Capture Simulation Initiative (CCSI) (Miller *et al.*, 2014). The open source CFD code Multiphase Flow with Interphase eXchanges (MFIX) (Benyahia *et al.*, 2012) was used as the simulator of the bubbling fluidized bed. The experimental setup and the MFIX model used to simulate it are fully documented in Lai *et al.* (2014). Below, only an abridged description of the data is provided.

The variables involved in the experimental data are the input variables flow rate (x_1 , **FRate**) and bed temperature (x_2 , **Temp**), and the output variable (y) is the pressure drop at location P3820 (i.e., the pressure drop across the bubbling fluidized bed). The pressure drop output is the time averaged value of the pressure drop once it was oscillating in steady state. The P3820 pressure drop was measured on the physical C2U system at a design of 44 distinct input settings. A total

of 60 MFI simulation cases were also designed and run for the purpose of emulator estimation. Both data sets are available online at the journal website. The MFI model parameters involved in the calibration for this case were Res-PP (θ_1): the particle-particle coefficient of restitution, Res-PW (θ_2): the particle-wall coefficient of restitution, FricAng-PP (θ_3): the particle-particle friction angle, FricAng-PW (θ_4): the particle-wall friction angle, PBVF (θ_5): Packed bed void fraction, and Part-Size (θ_6): the effective particle diameter of the sorbent material. The allowable ranges of the model parameters were chosen to be the same as those devised in Lai *et al.* (2014), mostly from literature review.

Table 1 provides the estimated θ values along with 95% confidence intervals and credible intervals, respectively, for the proposed **fboot-rs** method and the **bss-anova** method, both are numerically tested by simulations described in Section S2 of the supplemental document:

1. **fboot-rs**: The proposed frequentist method coupled with the bootstrap procedure of Section 3 *with* re-sampling of the design (i.e., keep Step 1).
2. **bss-anova**: Calibration of computational models via Bayesian smoothing spline ANOVA (Storlie *et al.*, 2014).

Both methods largely agree on their respective estimates and CIs, which provides some confirmation of the result. The first five parameters have fairly wide CIs relative to their allowable ranges, indicating that most of the range of these parameters produces reasonable model results. However θ_6 (effective particle size) does have tighter CIs and it seems as though values closer to 117 are preferred.

Figure 1 provides a visual summary for the simulator fit to the experimental data along with confidence bands (accounting for uncertainty in the emulator and the value of θ). The simulator with discrepancy (i.e., reality) predictions are provided as well. The pressure drop (y) is plotted against **Temp** (x_2) for six distinct values of **FRate** (x_1). The experimental data is also provided (along with the estimated 2σ measurement error bars). The experimental data was binned into the closest value of the six displayed **FRate** for display purposes. It is clear that the discrepancy is trending upwards as **FRate** increases. Figure 2 makes this relationship more explicit by isolating

Table 1: Estimates of θ along with 95% Confidence Intervals and Credible Intervals, respectively, for the proposed **fboot-rs** method and **bss-anova**. The parameters were restricted the the ranges provided during the estimation procedures.

θ_0	Range	fboot-rs		bss-anova	
		$\hat{\theta}$	95% CI	$\hat{\theta}$	95% CI
$\theta_{0,1}$	[0.80, 1.00]	0.927	(0.828, 0.964)	0.908	(0.831, 0.978)
$\theta_{0,2}$	[0.80, 1.00]	0.831	(0.829, 0.969)	0.897	(0.823, 0.979)
$\theta_{0,3}$	[25, 45]	39.5	(26.1, 41.5)	31.1	(25.4, 39.8)
$\theta_{0,4}$	[25, 45]	33.4	(25.5, 39.3)	31.4	(25.3, 41.0)
$\theta_{0,5}$	[0.30, 0.40]	0.346	(0.316, 0.388)	0.349	(0.313, 0.386)
$\theta_{0,6}$	[99, 125]	117	(110, 120)	115	(108, 120)

the discrepancy main effect functions across **FRate** and **Temp**, respectively. While it is evident that there is some statistically significant model discrepancy here, such discrepancy is relatively small when considering the magnitude of the pressure drop: the discrepancy is on the order of ~ 0.05 kPa, while the pressure drop is on the order of 0.72 kPa, thus a relative error of roughly 7%. Thus, for practical purposes MFIX can be used for prediction of a bubbling fluidized bed, knowing that the model form discrepancy is negligible.

6 Concluding Remarks

In this work, we have provided a frequentist framework for computer model calibration. This framework applies a general semi-parametric data model with an emulator for expensive simulators and a discrepancy function, which allows discrepancy between the simulator and the physical reality. Despite the flexibility of the model, our proposed framework gives identifiable parametrizations for both the model parameters and the discrepancy function. A practical estimation procedure and theoretical guarantees are provided for the proposed framework. Finally, a bootstrapping approach to provide uncertainty quantification for virtually any quantity of interest has also been developed. Due to the simplicity of the proposed calibration framework and the corresponding bootstrap, this approach can be easily coupled with a variety of optimization methods and/or emulators, which is beneficial to practitioners.

Three objectives of calibration are identified by Brynjarsdóttir and O’Hagan (2014): to study

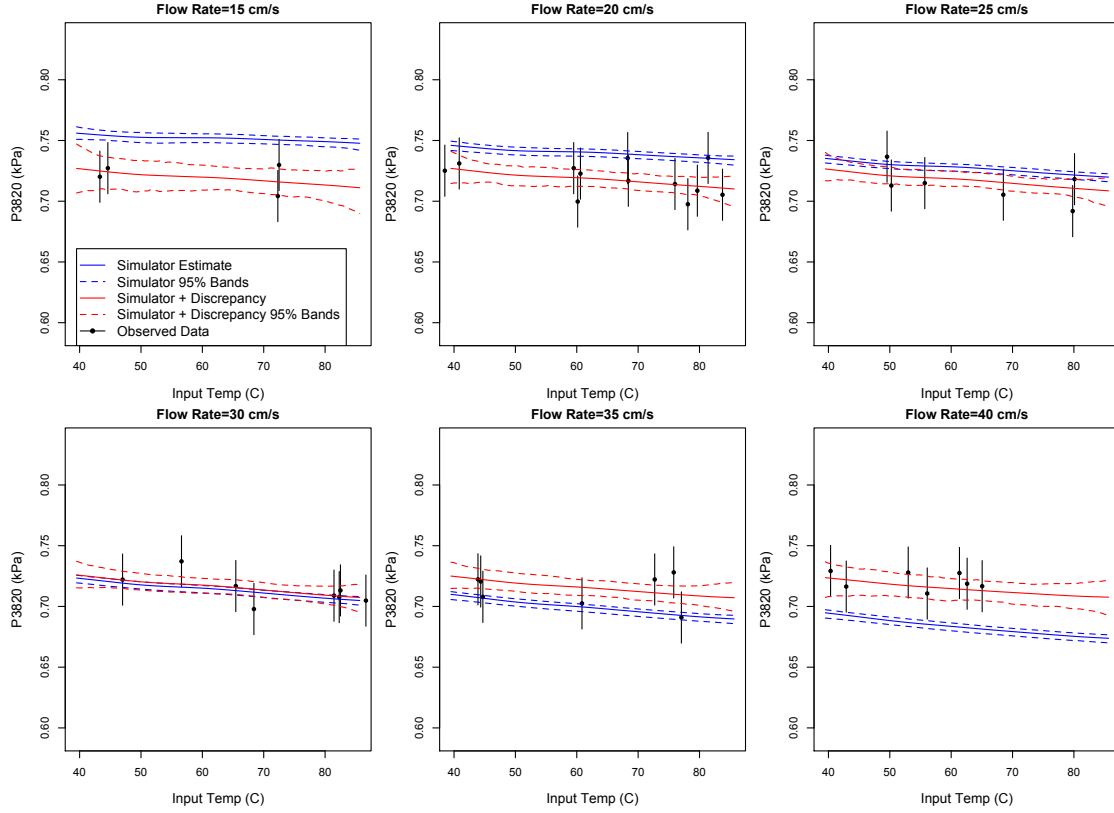


Figure 1: Fitted simulator (emulator) and simulator plus discrepancy curves (along with 95% confidence bands) as a function of Temperature at six distinct Flow Rates. Experimental data is binned into the closest value of the displayed Flow Rates for display.

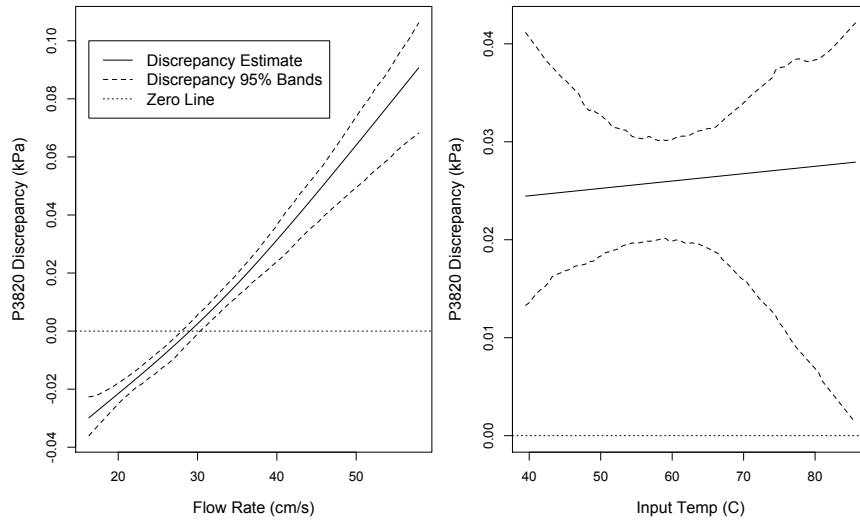


Figure 2: P3820 Discrepancy main effects across Flow Rate and Temperature.

the true values of the physical parameters, to predict the physical system’s behaviour within the scope of the observed data (i.e., interpolation) and outside the scope of the data (i.e., extrapolation). It is important to examine the usefulness of the proposed approach with respect to these objectives.

The proposed approach was not designed to, in general, address the first objective, unless the discrepancy δ is negligible and θ represents the physical parameters. The proposed approach, however, should work well for the second interpolation objective, although at times it may not be clear what is the physical meaning of θ . The third objective (extrapolation) is much more challenging. As pointed out by Brynjarsdóttir and O’Hagan (2014), the key to successful extrapolation requires that both θ and δ be meaningfully defined, and that accurate prior information for them are available. Under these situations Bayesian approach would be useful, and it is not advisable to use the proposed approach, especially when δ is large.

This paper is not suggesting that the proposed approach is correct for all problems. However, when there is no good prior information about some of the model parameters and/or discrepancy, other approaches may not be applicable or could produce unreliable results. It is in such cases that the proposed approach can be beneficial above and beyond the general benefits provided by the analytic study of its properties.

References

- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H. and Tu, J. (2007) A framework for validation of computer models. *Technometrics*, **49**, 138–154.
- Benyahia, S., Syamlal, M. and O’Brien, T. (2012) Summary of MFIx equations 2012-1. *National Energy Technology Laboratory Technical Report*. URL <https://mf1x.net1.doe.gov/documentation/MFIxEquations2012-1.pdf>.
- Brynjarsdóttir, J. and O’Hagan, A. (2014) Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, **30**, 114007.
- Conti, S., Gosling, J. P., Oakley, J. and O’Hagan, A. (2009) Gaussian process emulation of dynamic computer codes. *Biometrika*, **96**, 663–676.

- Curran, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1991) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, **86**, 953–963.
- Drignei, D. and Morris, M. D. (2006) Empirical Bayesian analysis for computer experiments involving finite-difference codes. *Journal of the American Statistical Association*, **101**, 1527–1536.
- Efron, B. and Tibshirani, R. J. (1993) *An introduction to the bootstrap*. New York: Chapman & Hall.
- Fang, K.-T., Li, R. and Sudjianto, A. (2010) *Design and modeling for computer experiments*. Boca Raton: Chapman & Hall.
- Hall, P. (1992a) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, **20**, 675–694.
- Hall, P. (1992b) On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics*, **20**, 695–711.
- Härdle, W. and Bowman, A. W. (1988) Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, **83**, 102–110.
- Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008) Computer model validation using high-dimensional output. *Journal of the American Statistical Association*, **103**, 570–583.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafo, J. A. and Ryne, R. D. (2004) Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, **26**, 448–466.
- Huang, D., Allen, T. T., Notz, W. I. and Zeng, N. (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, **34**, 441–466.
- Jones, D. R., Schonlau, M. and Welch, W. J. (1998) Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, **13**, 455–492.

- Joseph, V. R. and Melkote, S. N. (2009) Statistical adjustments to engineering models. *Journal of Quality Technology*, **41**, 362–375.
- Kennedy, M. C. and O’Hagan, A. (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, **63**, 425–464.
- Lai, K., Xu, Z., Pan, W., Shadle, L., Storlie, C., Dietiker, J., Li, T., Dartevelle, S. and Sun, X. (2014) Hierarchical calibration and validation of high-fidelity CFD models with C2U experiments. *Milestone Report*.
- Lane, W., Storlie, C., Montgomery, C. and Ryan, E. (2014) Numerical modeling and uncertainty quantification of a bubbling fluidized bed with immersed horizontal tubes. *Powder Technology*, **253**, 733–743.
- Miller, D., Syamlal, M., Mebane, D., Storlie, C., Bhattacharyya, D., Sahinidis, N., Agarwal, D., Tong, C., Zitney, S., Sarkar, A., Sun, X., Sundaresan, S., Ryan, E., Engel, D., and Dale, C. (2014) Carbon capture simulation initiative: A case study in multi-scale modeling and new challenges. *Annual Review of Chemical and Biomolecular Engineering*, **5**, 301–323.
- Oakley, J. and O’Hagan, A. (2004) Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society: Series B*, **66**, 751–769.
- Reich, B. J., Storlie, C. B. and Bondell, H. D. (2009) Variable selection in Bayesian smoothing spline ANOVA models: Application to deterministic computer codes. *Technometrics*, **51**, 110–120.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. New York: Cambridge University Press.
- Santner, T., Williams, B. and Notz, W. (2003) *The Design and Analysis of Computer Experiments*. New York: Springer.
- Sekhon, J. S. and Mebane, W. R. (1998) Genetic optimization using derivatives. *Political Analysis*, **7**, 187–210.

- Storlie, C. B., Lane, W. A., Ryan, E. M., Gattiker, J. R. and Higdon, D. M. (2014) Calibration of computational models with categorical parameters and correlated outputs via Bayesian smoothing spline ANOVA. *Journal of the American Statistical Association*. To appear.
- Storlie, C. B., Reich, B. J., Helton, J. C., Swiler, L. P. and Sallaberry, C. J. (2013) Analysis of computationally demanding models with continuous and categorical inputs. *Reliability Engineering & System Safety*, **113**, 30–41.
- Storlie, C. B., Swiler, L. P., Helton, J. C. and Sallaberry, C. J. (2009) Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, **94**, 1735–1763.
- van de Geer, S. (2000) *Empirical Processes in M-estimation*. New York: Cambridge University Press.
- Vecchia, A. V. and Cooley, R. L. (1987) Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research*, **23**, 1237–1250.
- Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: SIAM.
- Walker, S. G. (2013) Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*, **143**, 1621–1633.